Available online at www.sciencedirect.com

ScienceDirect

IFAC Papers Online CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 51-2 (2018) 765–770

Adaptive optimal operation of a parallel robotic liquid handling station ⋆

Tilman Barz * **Andreas Sommer** ** **Terrance Wilms** ***
Peter Neubauer *** **M. Nicolas Cruz Bournazou** ***

* *AIT Austrian Institute of Technology GmbH, Center for Energy, Vienna, Austria (e-mail: tilman.barz@ait.ac.at).*
** *Interdisciplinary Center for Scientific Computing (IWR), Heidelberg University, Heidelberg, Germany*
*** *Chair of Bioprocess Engineering, Institute of Biotechnology, Technische Universität Berlin, Berlin, Germany*

Abstract:
Results are presented from the optimal operation of a fully automated robotic liquid handling station where parallel experiments are performed for calibrating a kinetic fermentation model. To increase the robustness against uncertainties and/or wrong assumptions about the parameter values, an iterative calibration and experiment design approach is adopted. Its implementation yields a stepwise reduction of parameter uncertainties together with an adaptive redesign of reactor feeding strategies whenever new measurement information is available. The case study considers the adaptive optimal design of 4 parallel fed-batch strategies implemented in 8 mini-bioreactors. Details are given on the size and complexity of the problem and the challenges related to calibration of over-parameterized models and scarce and non-informative measurement data. It is shown how methods for parameter identifiability analysis and numerical regularization can be used for monitoring the progress of the experimental campaigns in terms of generated information regarding parameters and selection of the best fitting parameter subset.

Keywords: Parallel robotic liquid handling station, *E. coli* kinetic model, Optimal experimental design for model calibration, Adaptive input design, Identifiability and ill-conditioning analysis

1. INTRODUCTION

The development from product to manufacturing is known to be the bottleneck in the bio-industry (Neubauer et al. (2013)). Hence there is a strong interest in biotechnology to accelerate, systematize, and increase the reliability of bioprocess development. Recently, the advances in miniaturization, speed, and parallelization of experiments has set the path for a faster and cheaper generation of relevant process information to support consistent bioprocess development. Liquid Handling Stations (LHS) for High Throughput Screening (HTS) and High Throughput Bioprocess Development (HTBD) have been created exploiting the fast development in robotics, automation, and sensor technology (Wiendahl et al. (2008)).

The challenge now is to design experimental campaigns that exploit the capabilities of these facilities in an optimal way. Although there is a community devoted to research in this direction (see e.g. Unthan et al. (2015); Nickel et al. (2017)), most approaches work only for single timepoint (static) experiments and do not consider the dynamic evolution of biological systems properly. This is a relevant issue, not only because microorganisms have a very complex dynamical interaction with the environment, but also because most industrial processes run in highly nonlinear conditions. It is generally recognized that kinetic (dynamic) modeling and model based analysis are in principle of very high value for industrial biotechnology supporting the rational design of cell factory properties and the design of the bioreactor or fermentation process (Almquist et al. (2014)). However, several challenges remain before kinetic modeling will reach the degree of maturity required for routine application in industry, see Almquist et al. (2014).

This contribution focuses on systematic approaches for planning, execution and analysis of parallel experiments for kinetic model development purposes, making optimal use of robotic liquid handling stations. For doing so, Optimal Experimental Design (OED) for model calibration is used to maximize the information content in measured data generated by process analytics and online sensors during a fed-batch running experiment, see e.g. Versyck et al. (1997). The OED seeks experimental feeding strategies that minimize the uncertainty of the model parameters recovered from Parameter Estimation (PE). These optimally designed experiments can noticeable reduce the experimental effort compared to conventionally or heuristically designed ones (Bauer et al. (2000)).

⋆ T.B. and A.S. acknowledge partial funding of this project by the Austrian Research Funding Association (FFG) within the programme Bridge in the project modELTES (project No. 851262). M.N.C.B. acknowledge financial support by the German Federal Ministry of Education and Research (BMBF) within the Framework Concept 'Research for Tomorrow's Production' (AUTOBIO).

10.1016/j.ifacol.2018.04.006

Typical kinetic models are nonlinear in the parameters and the experimental design is sensitive to parametric uncertainties. To perform a design, an initial guess of the parameters, i.e. the best available parameter values, is used. Thus, the quality of a computed optimal design depends on the quality of this parameter guess (Bauer et al. (2000)). A widely used approach to cope with uncertainties in the parameter estimates is based on the iterative refinement of the experimental design whenever new measurements and parameter estimates are available. This means that experiments are designed, executed and analyzed in a sequence (Bauer et al. (2000)). This stepwise reduction of the parameter uncertainty leads to more reliable model predictions and designs that are closer to a truly optimal experiment. The most efficient implementation of this approach for dynamic systems is the online or adaptive experimental redesign, an idea which was already discussed in the early 70ies, see Mehra (1974). Here the experiment is iteratively re-designed and parameters are re-estimated as information is generated. By this, it is possible to exploit new measurement information as soon as it is generated by the running experiment minimizing the mismatch between calculated and real outputs.

When multiple equipment pieces are available, running parallel experiments is highly advantageous in terms of time and use of resources. Unfortunately, when planning a large number of experiments in parallel, the number of designs with uncertain parameters is also large (the studied experimental platform can operate up to 48 reactors in parallel). In the worst case, in parallel settings all experiments are computed with very poor initial guesses, where even the order of magnitude might be wrong. This makes the quality of the information in the generated data for parameter identification highly unreliable. Moreover, overparameterized models and scarce informative experimental data in quantity as well as in quality additionally present serious challenges, especially in online or real-time applications with recursive estimation and optimization settings (Barz et al. (2016)). Corresponding ill-posed problems are highly problematic as they might destabilize the solution and affect the reliability of the estimates. Furthermore, in the presence of ill-posed PE problems a redesign of optimal inputs (for parameter precision improvement) leads to ineffective and/or meaningless designs (López Cárdenas et al. (2015)).

The paper is organized as follows. First, the mathematical problem of the adaptive experimental redesign for calibration of a kinetic *E. coli* fermentation model in a parallel robotic liquid handling station is presented. It is discussed how to integrate an identifiability and ill-conditioning analysis in the framework. The fed-batch fermentation case study is briefly presented and information on the dimensionality of the mathematical problem is given. Experimental results from the adaptive optimal robot operation are presented. Focus is on the numerical condition of the re-estimation and redesign problems and the implications for parameter identifiability during the experimental run. Finally, conclusions and directions for further research are given.

2. MATHEMATICAL PROBLEM FORMULATION

Kinetic models in industrial biotechnology are formulated as systems of nonlinear differential equations (Almquist et al. (2014)). For each reactor $r \in \mathcal{R}$ the model reads:

$$\left.\begin{aligned} \dot{x}_r(t) &= f\left(x_r(t), u_r(t), \theta\right) \\ y_r(t) &= \mathrm{A}\, x_r(t) \\ x_r(t_0) &= x_{0,r} \end{aligned}\right\} \forall r \in \mathcal{R} \tag{1}$$

where $t \in [t_0, t_{end}] \subseteq \mathbb{R}$ is the time, $x_r(t) \in \mathbb{R}^{n_x}$ are dependent state variables, $u_r(t) \in \mathbb{R}^{n_u}$ are the time-varying inputs (or experimental design variables) and $\theta \in \mathbb{R}^{n_\theta}$ the unknown parameter vector and initial conditions are given by $x_{0,r}$. The vector $y_r(t) \in \mathbb{R}^{n_y}$ are the predicted response variables (for which sensors are available) whose elements are defined by the selection matrix $A \in \mathbb{R}^{n_y \times n_x}$. For partially observed models $n_y < n_x$ (not all states are measured).

The experimental setup of the parallel robotic platform consists of

- n_r parallel experiments $\mathcal{R} := \{1, \cdots, n_r\}$, which are executed in $r \in \mathcal{R}$ reactors numbered from 1 to n_r.

The robot simultaneously feeds all reactors $\mathcal{R}$ at

- n_f discrete feeding times $\mathcal{F} := \{\mathrm{T}_1, \cdots, \mathrm{T}_{n_f}\}$, where n_f is the total number of feeds(=injections) into one reactor and $t_0 \leq \mathrm{T}_i \leq t_{end}$.

The robot continuously monitors/takes samples from all reactors $\mathcal{R}$ at

- n_s discrete measurement times $\mathcal{M} := \{\tau_1, \cdots, \tau_{n_s}\}$, where n_f is the total number of recorded measurements from one reactor [1] and $t_0 \leq \tau_j \leq t_{end}$.

Each reactor $r \in \mathcal{R}$ has its individual feeding strategy, defined by the discrete inputs $u_r(\mathrm{T}_i) \in \mathbb{R}^{n_u}$ with $\mathrm{T}_i \in \mathcal{F}$. The discrete inputs represent injections and are collected in the vector:

$$\mathrm{u}_r = \begin{bmatrix} u_r(\mathrm{T}_1) \\ \vdots \\ u_r(\mathrm{T}_{n_f}) \end{bmatrix} \in \mathbb{R}^{n_u \cdot n_f} \tag{2}$$

with the number of individual species to be injected n_u and the total number of injections n_f.
From each reactor $r \in \mathcal{R}$ measurements $y^m_{r,j} \in \mathbb{R}^{n_y}$ are recorded. These discrete measurements are obtained from continuous and/or at-line concentration analysis at time instances $\tau_j \in \mathcal{M}$. They are collected in the vector:

$$\mathrm{y}^m_r = \begin{bmatrix} y^m_{r,1} \\ \vdots \\ y^m_{r,n_s} \end{bmatrix} \in \mathbb{R}^{n_y \cdot n_s} \tag{3}$$

with the number of measured states n_y and the number of samplings taken by the robot n_s. Corresponding predicted response variables $y_r(\tau_j, \mathrm{u}_r, \theta) \in \mathbb{R}^{n_y}$ (simulated measurements) are evaluated at the same time instances $\tau_j \in \mathcal{M}$. They are collected in the vector

[1] Note that, for the sake of simplicity, it is assumed that all measurements are recorded at same/identical measurement times. In practice oxygen is monitored continuously (high frequency measurements) whilst concentrations are analysed from liquid samples taken by the robot (low frequency measurements).

$$\mathrm{y}_r(\mathrm{u}_r,\theta)=\begin{bmatrix} y_r(\tau_1,\mathrm{u}_r,\theta)\\ \vdots \\ y_r(\tau_{n_s},\mathrm{u}_r,\theta)\end{bmatrix}\in\mathbb{R}^{n_y\cdot n_s} \tag{4}$$

The predicted responses are obtained from the solution of (1) for each reactor $r\in\mathcal{R}$ and therefore depend on the corresponding discrete inputs u_r and the parameters θ.

In the following $\mathrm{u}_\mathcal{R}$ is used to combine all discrete inputs of all reactors $\mathcal{R}$ in one vector.

$$\mathrm{u}_\mathcal{R}=[\mathrm{u}_1^T,\cdots,\mathrm{u}_{n_r}^T]^T\in\mathbb{R}^{n_r\cdot n_u\cdot n_f} \tag{5}$$

2.1 Parameter estimation and experimental design

We assume that the deterministic model in (1) is an exact structural model and that corresponding discrete inputs u_r and initial state variables $x_{0,r}$ are known. Measurements (taken from each reactor $r\in\mathcal{R}$ at measurement time instant $\tau_j\in\mathcal{M}$) are assumed to be normally distributed with zero mean and known covariance matrix $\mathrm{V}\in\mathbb{R}^{n_y\times n_y}$. Correlations between errors in different reactors and autocorrelation (or serial correlation) are not considered. With the assumptions above, the weighted least squares objective delivers a maximum-likelihood estimate of the parameters. Accordingly, parameter estimates $\hat\theta$ are obtained from the solution of the unconstrained problem Bard (1974):

$$\hat\theta=\arg\min_\theta \phi(\mathrm{u}_\mathcal{R},\theta) \tag{6}$$

The objective function in (6) is defined by weighted least squares with the weights given by the inverse of V:

$$\phi(\mathrm{u}_\mathcal{R},\theta)=\frac{1}{2}\sum_{r\in\mathcal{R}}\sum_{\tau_j\in\mathcal{M}} e_r(\tau_j,\mathrm{u}_r,\theta)\cdot\mathrm{V}^{-1}\cdot\left(e_r(\tau_j,\mathrm{u}_r,\theta)\right)^T \tag{7}$$

with the residual vector defined as

$$e_r(\tau_j,\mathrm{u}_r,\theta)=y_r(\tau_j,\mathrm{u}_r,\theta)-y^m_{r,j} \tag{8}$$

If the estimate $\hat\theta$ is the unconstrained minimum of (6), then the covariance matrix of the estimates can be approximated by Bard (1974):

$$\mathrm{C}(\mathrm{u}_\mathcal{R},\hat\theta)\cong\left(\sum_{r\in\mathcal{R}}\sum_{\tau_j\in\mathcal{M}} {S_r}^T(\tau_j,\mathrm{u}_r,\hat\theta)\cdot\mathrm{V}^{-1}\cdot S_r(\tau_j,\mathrm{u}_r,\hat\theta)\right)^{-1} \tag{9}$$

where $S_r(\tau_j,\mathrm{u}_r,\hat\theta)$ is the sensitivity matrix, with

$$S_r(\tau_j,\mathrm{u}_r,\hat\theta)=-\frac{\partial e_r(\tau_j,\mathrm{u}_r,\hat\theta)}{\partial\theta}\in\mathbb{R}^{n_y\times n_\theta} \tag{10}$$

$\mathrm{C}(\mathrm{u}_\mathcal{R},\hat\theta)$ gives information on the precision of the maximum-likelihood estimate $\hat\theta$. Designing an optimal experiment for improving parameter precision means to minimize some criterion on this matrix by optimally choosing the discrete inputs $\mathrm{u}_\mathcal{R}$ (Bauer et al. (2000)):

$$\mathrm{u}^*_\mathcal{R}:=\arg\min_{\mathrm{u}_\mathcal{R}}\psi^A(\mathrm{u}_\mathcal{R},\hat\theta) \tag{11}$$

A commonly used metric is the so called A-optimal criterion ψ^A (the trace), which yields the optimal experimental design objective function:

$$\psi^A(\mathrm{u}_\mathcal{R},\hat\theta):=\frac{1}{n_\theta}\mathrm{Tr}\left[\mathrm{C}(\mathrm{u}_\mathcal{R},\hat\theta)\right] \tag{12}$$

2.2 Sequential re-estimation and experiment redesign

The experimental design criterion ψ^A is a (nonlinear) function of the current parameter estimates $\hat\theta$ in (6). Wrong assumptions or outdated parameter estimates may therefore severely affect the quality of an experimental design and the OED criterion ψ^A may deteriorate. This is the case if $\hat\theta$ differs from the initial (or prior) estimate for which an optimal experiment was designed and realized. To increase the robustness against parameter uncertainties, the OED method is implemented as adaptive experimental redesign. [2] This iterative strategy takes full advantage of new available information on the parameter values.

Consider additional

- n_f^+ feeding times $\mathcal{F}^+:=\{\mathrm{T}_{n_f+1},\cdots,\mathrm{T}_{n_f+n_f^+}\}$, with corresponding feeds $u_r^+(\mathrm{T}_i)$, with $\mathrm{T}_i\in\mathcal{F}^+$ and $r\in\mathcal{R}$,
- n_s^+ measurement times $\mathcal{M}^+:=\{\tau_{n_s+1},\cdots,\tau_{n_s+n_s^+}\}$, and corresponding measurements $y^m_{r,j}\in\mathbb{R}^{n_y}$, with $\tau_j\in\mathcal{M}^+$ and $r\in\mathcal{R}$.

In the same way as in (2) and (5), the additional discrete inputs are collected as $\mathrm{u}_r^+\in\mathbb{R}^{n_u\cdot n_f^+}$, and combined:

$$\mathrm{u}^+_\mathcal{R}=[\mathrm{u}_1^{+T},\cdots,\mathrm{u}_{n_r}^{+T}]^T\in\mathbb{R}^{n_r\cdot n_u\cdot n_f^+} \tag{13}$$

The current state of knowledge concerning the parameter values results from the prior estimation, i.e. solution of (6). The objective function of the re-estimation problem for $(n_s+n_s^+)$ measurement time instances reads:

$$\phi^{upd}(\mathrm{u}^+_\mathcal{R},\theta)=\phi(\theta)+\phi^+(\mathrm{u}^+_\mathcal{R},\theta) \tag{14}$$

$\phi^+(\mathrm{u}^+_\mathcal{R},\theta)$ contains the additional $n_y\cdot n_s^+$ residuals:

$$e_r^+(\tau_j,\mathrm{u}^+_\mathcal{R},\theta)=y_r^+(\tau_j,\mathrm{u}^+_\mathcal{R},\theta)-y^{m+}_{r,j}\quad\text{with}\quad r\in\mathcal{R},\ \tau_j\in\mathcal{M}^+ \tag{15}$$

Note that in (14) and (15) there are no dependencies on past $\mathrm{u}_\mathcal{R}$, these past inputs have already been realized and therefore are considered as constant. The solution to the re-estimation problem updates the prior estimate $\hat\theta$ with the posterior estimate $\hat\theta^{upd}$. The covariance matrix of $\hat\theta^{upd}$ can be approximated by:

$$\mathrm{C}^+(\mathrm{u}^+_\mathcal{R},\hat\theta^{upd})\cong\Bigg(\left(\mathrm{C}(\hat\theta^{upd})\right)^{-1} + \sum_{r\in\mathcal{R}}\sum_{\tau_j\in\mathcal{M}^+} {S_r}^T(\tau_j,\mathrm{u}^+_\mathcal{R},\hat\theta^{upd})\cdot\mathrm{V}^{-1}\cdot S_r(\tau_j,\mathrm{u}^+_\mathcal{R},\hat\theta^{upd})\Bigg)^{-1} \tag{16}$$

where the first term represents information from the prior measurements and the second term represents information gained by the additional future measurements.

The optimal redesign of experiments computes future inputs $\mathrm{u}^+_\mathcal{R}$ by maximizing the information content in future measurements taken at $\tau_j\in\mathcal{M}^+$ whenever new parameter estimates are available. However, since these future measurements have not been taken yet, $\hat\theta^{upd}$ is not known. Thus, in order to optimize these future measurements, the

[2] The parameters are sequentially re-estimated and the feeding strategy is adaptively refined (i.e. future input variables are re-optimized) during the course of an experiment. The frequency of these re-estimations and re-optimizations depends on the complexity of the involved numeric computations and the systems dynamics and sampling rates, i.e. the availability of new informative measurements.

redesign problem is solved for the prior estimate $\hat{\theta}$ in place of $\hat{\theta}^{upd}$ (Bard (1974)). Applying the A-optimal criterion to the covariance matrix in (16), the objective function of the experiment redesign problem reads:

$$\psi^A(\mathrm{u}_{\mathcal{R}}^+, \hat{\theta}) = \frac{1}{n_\theta} \mathrm{Tr}\left[\mathrm{C}^+(\mathrm{u}_{\mathcal{R}}^+, \hat{\theta})\right] \tag{17}$$

The design variables $\mathrm{u}_{\mathcal{R}}^+$ influence future measurements at sampling time instances $\mathcal{M}^+$ only, and thus the second term in (16). Thus, the first term in (16) is a constant matrix. As soon as the new measurements are taken, parameters are re-estimated and the iteration repeats/continues with the updated parameter estimate in the role of $\hat{\theta}$.

2.3 Identifiability and ill-conditioning analysis

The analysis is performed locally based on the sensitivity matrix (10), which contains useful information for numerical analysis of the PE and OED problem (López Cárdenas et al. (2015)): In derivative-based solution methods for PE, the sensitivity matrix $\tilde{\mathrm{S}}_{\mathcal{R}}$, see (18), is used for the calculation of the step direction, i.e. computation of the Jacobian and Hessian. Thus, the condition of $\tilde{\mathrm{S}}_{\mathcal{R}}$ characterizes the condition of the PE problem. Moreover, there exists a direct relationship between the singular values of $\tilde{\mathrm{S}}_{\mathcal{R}}$ and the commonly used criteria for parameter identifiability and OED, see (11), namely the eigenvalues of the Hessian and of the parameter covariance matrix, see (9).

$\tilde{\mathrm{S}}_{\mathcal{R}}$ is scaled and is updated (quasi continuously) for every new measurement and the current parameter estimate $\hat{\theta}$:

$$\tilde{S}_r(\tau_j, \mathrm{u}_r, \hat{\theta}) = \left(\mathrm{V}^{\frac{1}{2}}\right)^{-1} S_r(\tau_j, \mathrm{u}_r, \hat{\theta}) \tag{18}$$

where the weighting matrix $\mathrm{V}^{\frac{1}{2}} \in \mathbb{R}^{n_y \times n_y}$ is the square root of the measurement covariance matrix V (compare with (7)). Note that it is assumed that parameter values are normalized by their respective initial guesses. Sensitivities corresponding to all discrete measurements of one reactor $r \in \mathcal{R}$ are collected as:

$$\tilde{\mathrm{S}}_r(\mathrm{u}_r, \theta) = \begin{bmatrix} \tilde{S}_r(\tau_1, \mathrm{u}_r, \theta) \\ \vdots \\ \tilde{S}_r(\tau_{n_s}, \mathrm{u}_r, \theta) \end{bmatrix} \in \mathbb{R}^{(n_y \cdot n_s) \times n_\theta} \tag{19}$$

and combining the results for all reactors $\mathcal{R}$ gives:

$$\tilde{\mathrm{S}}_{\mathcal{R}}(\mathrm{u}_{\mathcal{R}}, \theta) = \begin{bmatrix} \tilde{\mathrm{S}}_1(\mathrm{u}_1, \theta) \\ \vdots \\ \tilde{\mathrm{S}}_{n_r}(\mathrm{u}_{n_r}, \theta) \end{bmatrix} \in \mathbb{R}^{(n_r \cdot n_y \cdot n_s) \times n_\theta} \tag{20}$$

The numerical problem analysis is done by computation of the singular values (SVs) of $\tilde{\mathrm{S}}_{\mathcal{R}}$. Note that the computation of the SVs must/can be done numerically stable also for ill-posed PE and OED problems. Indicators for the identifiability and ill-conditioning analysis are derived from the analysis of the singular value spectrum, condition number, and collinearity index of $\tilde{\mathrm{S}}_{\mathcal{R}}$. A rank-revealing singular value decomposition (SVD) is computed:

$$\tilde{\mathrm{S}}_{\mathcal{R}} = \sum_{i=1}^{n_\theta} \mu_i \varsigma_i v_i^T \tag{21}$$

with ς_i being the i-th singular value of $\tilde{\mathrm{S}}_{\mathcal{R}}$ (ordered according to magnitude as $\varsigma_1 \geq \varsigma_2 \geq \cdots \geq \varsigma_{n_\theta} \geq 0$), and the left and right singular vectors $\mu_i \in \mathbb{R}^{n_r \cdot n_y \cdot n_s}$ and $v_i \in \mathbb{R}^{n_\theta}$, respectively. The number of linearly independent parameters corresponds to the numerical ε-rank r_ε of $\tilde{\mathrm{S}}_{\mathcal{R}}$. r_ε is defined by the maximum number of ς_i with $i = 1, \cdots, n_\theta$ for which the sub-condition number $\kappa_i = \varsigma_1/\varsigma_i$ and the sub-collinearity index $\gamma_i = 1/\varsigma_i$ are below a critical threshold. Corresponding upper bounds, namely the maximum condition number (large ratio in SVs) and the maximum collinearity index (smallness in SVs) are defined by empirical values, $\kappa^{max} = 1000$ and $\gamma^{max} = 10^{10} \ldots 10^{15}$, respectively.[3] Accordingly, κ^{max} assures numerical stability and γ^{max} controls linear dependencies.

A useful graphic representation is a plot of the SVs against their index i. This gives the so called SV spectrum. For an ill-posed problem this spectrum can be partitioned in the first SVs which define a well posed problem (their number equals r_ε by definition) and the remaining ill-conditioned SVs. Accordingly, well-conditioned SVs are above a lower bound, namely the ε-threshold, which is defined by

$$\varepsilon = \max\left\{\varepsilon_\kappa = \frac{\varsigma_1}{\kappa^{max}(\tilde{\mathrm{S}}_{\mathcal{R}})}, \quad \varepsilon_\gamma = \frac{1}{\gamma^{max}(\tilde{\mathrm{S}}_{\mathcal{R}})}\right\} \tag{22}$$

A forward selection method using orthogonal projections of $\tilde{\mathrm{S}}_{\mathcal{R}}$ is used to assess parameter identifiability. This widely used approach for parameter subset selection seeks an ordering of parameters according to the linear independence of the columns of $\tilde{\mathrm{S}}_{\mathcal{R}}$ by applying QRP decomposition (see López Cárdenas et al. (2015)). The result is the selection of a well-conditioned parameter subset. It is here also used as adaptive regularization strategy, transforming ill-conditioned into well-conditioned (but reduced) PE and OED problems.

3. CASE STUDY

The case study considers the optimal operation of a parallel robotic liquid handling station (see fig. 1) for calibration of a macro-kinetic *E. coli* fermentation model. The reader is referred to Nickel et al. (2017) for details on the experimental facility and to Cruz Bournazou et al. (2017) for details on the modeling and optimal operation. The experimental facility was set to run 8 fed-batch cultivations in 42 parallel mini BioReactors with a total working volume of 9-14 mL. The strain used is the widely studied *Escherichia coli* W3110. This wild type is known for its fast replication

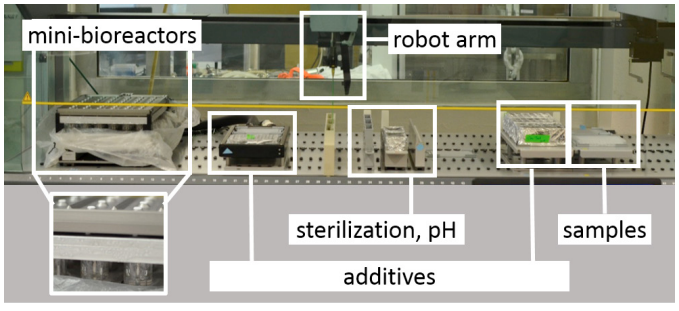

Fig. 1. Deck layout of the Freedom Evo LHS: bioreactor 48; reactor section, feeding section, samples, robot arm.

[3] It should be noted that while κ_i is scale invariant, γ_i is not. Moreover, it has been found that tuning of these maximum values might be useful in the particular application.

time of ≈ 60 min and high acetate production rate. Using the enzymatic glucose release system from EnBase^{TM}, the concept of glucose-limited fed-batch cultivation was emulated which is commonly used in industrial processes. Additionally, cycles of culture medium, enzyme, acetic acid, and glucose were added (feeding strategy, three times per hour) by the pipetting channel. The sampling was done every 20 minutes to obtain measurements of glucose, optical density, and acetate. The delay between sampling and analytics of the enzymatic assays was 80 minutes. The pH was controlled, dissolved oxygen and temperature were measured online (the data was recorded every 1 min).

Table 1. Problem characteristics.

Number of	Quantity	Comment
Reactors	$n_r = 4{\cdot}2$	Four reactors individually operated, each with a duplication.
Feeds	$n_u = 4$	Medium, enzyme, acetate, glucose.
Samples	$n_y = 1{+}3$	Oxygen; biomass, acetate, glucose. Frequency: 60/hour + 3/hour.
States	$n_x = 7$	Individual reactor model.
Parameters	$n_\theta = 25$	17 valid for all reactors, 1·8 specific for individual reactors.

A summary of the characteristic problem variables is given in table 1. The total length of the experimental run was 6 hours of operation, this makes a total number of

- 288 ($= 6 \cdot n_u \cdot n_f \cdot n_r/2$) experimental design variables, representing four individual feeding strategies;
- 3312 ($= 6 \cdot n_y \cdot n_s \cdot n_r$) measurement data points, representing data from all eight reactors;
- 56 ($= n_x \cdot n_r$) state variables/equations, which allows for individual analysis of reactor duplicates.

4. RESULTS

An exemplarily plot of the final OED and the final fitting of generated experimental data is depicted in Fig. 2.

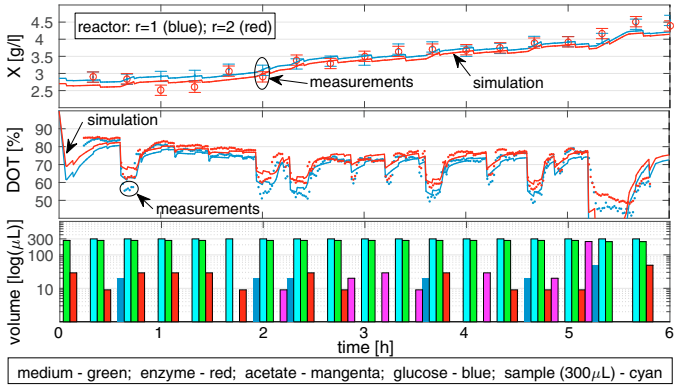

Fig. 2. Results from the adaptive optimal robot operation for reactors $r = 1$ and $r = 2$ (duplicates). First and second subplot show the fitting of the cell dry weight (X, in grams per liter) and the dissolved oxygen tension (DOT, in % of saturation). Third subplot shows realized feeding/ sampling strategy, where feed and extracted sample volumes are depicted with bars (semi-log graph in micro liters).

It turns out that compared to a heuristically chosen design, the realized OED generates data which reduces the average coefficient of variation of the parameter estimates by a factor of 50 considering all 23 (out of 25) identifiable parameters (Cruz Bournazou et al. (2017)).

Fig. 3 shows results from the identifiability and ill-conditioning analysis. The analysis is performed repeatedly to monitor changes in: the SV spectrum, the number of identifiable parameters, and the selected identifiable parameters. Note that logarithmic axes are used for time and SV. In Fig. 3 top, it can be seen that during the first half hour of the experiment the sensitivity matrix $\tilde{S}_{\mathcal{R}}$, see (21), has very low rank (maximum rank is 25 equal to n_θ). However, the rank is increasing fast and finally reaches 20. This behavior can be explained by the low number of measurements (at the very beginning the number of measurements is even smaller than the number of parameters) and low parameter sensitivities. The PE and OED problem can be classified as rank-deficient. The reduction of the parameter space can generate new better-conditioned problems. In Fig. 3 bottom, it is shown that up to the first parameter re-estimation, parameters are consecutively added to the identifiable subset (activated). For parameters 16 and 19 there are inconsistencies (activation, deactivation, activation) which result from shortcomings in the forward selection method. Note that the problem rank before and after the first parameter re-estimation drops from 21 to 16, for the second re-estimation it increases from 17 to 23. This unwanted behavior results from limitations by the local analysis of the nonlinear model. It presents a big problem as it can destabilize the iterative design and estimation approach, see Barz et al. (2016).

Fig. 4 shows a theoretical evolution of SV and parameter sets assuming that the finally identified parameters are known from the beginning. During the last last five hours the rank improves only by two, reaching finally 23. Here the problems are of ill-determined rank, due to insufficient measurement data and/or correlations in the parameters. Remarkably, the OED is not able to attract additional parameters to the identifiable region.

5. CONCLUSIONS AND OUTLOOK

The extensive requirements in biotechnology for experimental validation makes methods to compute and perform efficient experiments highly relevant. Adaptive methods for optimal design of parallel systems are definitely an important step towards fast and cheap calibration of kinetic models in bioengineering. The results of this work show that it is currently possible to fit kinetic models to experimental data while this is being generated. By this a validated model of a specific process is available before the end of the actual calibration experiment.

Interesting challenges are related to adaptive regularization strategies, which: support the iterative identification of a best-fitting identifiable parameter subset; improve the problem condition and attract unidentifiable parameters to the identifiable region; guarantee a numerically robust and stable iterative identification and experiment redesign; and, are applicable right from the start of the experimental campaign (i.e. are applicable for both, rank-deficient and ill-determined rank problems).

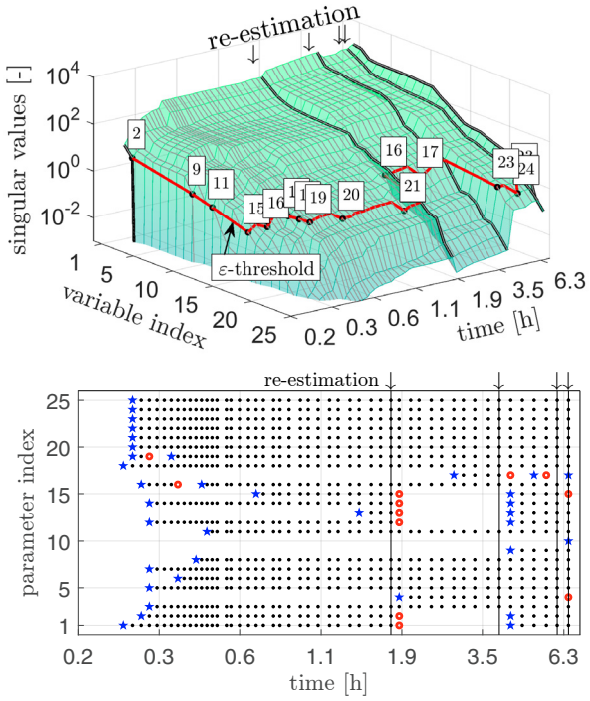

Fig. 3. Results from identifiability and ill-conditioning analysis during optimal robot operation. Time points where parameters are re-estimated (and feeding strategies are redesigned) are indicated by arrows.
Top: The evolution (in time) of the singular value spectrum $\{\varsigma_1, \cdots, \varsigma_{n_\theta}\}$ is severely affected by changing parameter values due to re-estimations. Bottom: Corresponding parameter set selection. Active parameters belong to the identifiable parameter subset. This subset changes significantly after parameter re-estimation 1 and 2.

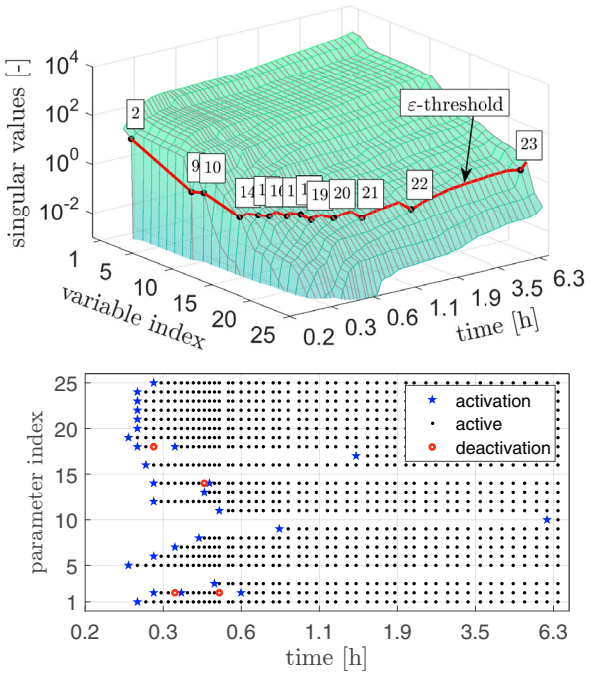

Fig. 4. Theoretical identifiability and ill-conditioning analysis for optimal robot operation. Parameter (re-)estimations are *not* performed. Instead, final parameter estimates are assumed to be known from the beginning.
Top: The evolution (in time) of the singular value spectrum $\{\varsigma_1, \cdots, \varsigma_{n_\theta}\}$ indicates continuous improvements in parameter identifiability and increase of the identifiable parameter subset. Bottom: Corresponding parameter set selection. Active parameters belong to the identifiable parameter subset. The analysis reveals only minor inconsistencies by the forward selection algorithm in the beginning of the experiment.

REFERENCES

Almquist, J., Cvijovic, M., Hatzimanikatis, V., Nielsen, J., and Jirstrand, M. (2014). Kinetic models in industrial biotechnology–improving cell factory performance. *Metabolic Engineering*, 24, 38–60.

Bard, Y. (1974). *Nonlinear parameter estimation.* Academic Press, New York.

Barz, T., López Cárdenas, D.C., Cruz Bournazou, M.N., Körkel, S., and Walter, S.F. (2016). Real-time adaptive input design for the determination of competitive adsorption isotherms in liquid chromatography. *Computers & Chemical Engineering*, 94, 104–116.

Bauer, I., Bock, H.G., Körkel, S., and Schlöder, J.P. (2000). Numerical methods for optimum experimental design in DAE systems. *Journal of Computational and Applied Mathematics*, 120(1), 1–25.

Cruz Bournazou, M., Barz, T., Nickel, D., López Cárdenas, D., Glauche, F., Knepper, A., and Neubauer, P. (2017). Online optimal experimental re-design in robotic parallel fedbatch cultivation facilities. *Biotechnology and Bioengineering*, 114(3), 610–619.

López Cárdenas, D.C., Barz, T., Körkel, S., and Wozny, G. (2015). Nonlinear ill-posed problem analysis in model-based parameter estimation and experimental design. *Computers & Chemical Engineering*, 77, 24–42.

Mehra, R. (1974). Optimal input signals for parameter estimation in dynamic systems–survey and new results. *IEEE Transact. on Automatic Control*, 19(6), 753–768.

Neubauer, P., Cruz Bournazou, N., Glauche, F., Junne, S., Knepper, A., and Raven, M. (2013). Consistent development of bioprocesses from microliter cultures to the industrial scale. *Engineering in Life Sciences*, 13(3), 224–238.

Nickel, D.B., Cruz-Bournazou, M.N., Wilms, T., Neubauer, P., and Knepper, A. (2017). Online bioprocess data generation, analysis, and optimization for parallel fed–batch fermentations in milliliter scale. *Engineering in Life Sciences*, 17(11), 1195–1201.

Unthan, S., Radek, A., Wiechert, W., Oldiges, M., and Noack, S. (2015). Bioprocess automation on a mini pilot plant enables fast quantitative microbial phenotyping. *Microbial Cell Factories*, 14(1), 32.

Versyck, K.J., Claes, J.E., and van Impe, J.F. (1997). Practical identification of unstructured growth kinetics by application of optimal experimental design. *Biotechnology Progress*, 13(5), 524–531.

Wiendahl, M., Schulze Wierling, P., Nielsen, J., Fomsgaard Christensen, D., Krarup, J., Staby, A., and Hubbuch, J. (2008). High throughput screening for the design and optimization of chromatographic processes—miniaturization, automation and parallelization of breakthrough and elution studies. *Chemical Engineering & Technology*, 31(6), 893–903.